

Distinguishability, Ensemble Steering, and the No-Signaling Principle

Joonwoo Bae¹

¹Center for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543 and, ICFO-Institut de Ciències Fotòniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain.

(Dated: October 12, 2012)

We consider a fundamental operational task, distinguishing systems in different states, in the framework of generalized probabilistic theories and provide a general formalism of the task. Optimality conditions are explicitly shown. With the formalism established, we show that for any generalized probabilistic theories where the ensemble steering is allowed, the no-signaling principle can tightly determine the distinguishability of states, independently to how their state spaces are structured. This shows that distinguishability is generally characterized as a consequence that ensemble steering on states and the non-signaling constraint on probabilities are compromised.

PACS numbers: 03.65.Ud, 03.67.Hk, 03.65.Wj

One of fundamental aspects of the quantum formalism is that both *states* and *probabilities* are needed to characterize given systems before and after measurement, respectively. This is connected to how quantum systems achieve practical advantages in information processing. For instance, nonlocal quantum probabilities [1] are resources for secure communication [2]. Interestingly, quantum systems cannot attain correlations that are maximally non-local in non-signaling frameworks [3]. Information-theoretic characterizations to quantum probabilities, such as the information causality as a way of generalizing the no-signaling condition [4] [5] [6], elucidate strengths and limitations of quantum systems, showing deep connections to quantum foundations and information applications.

On the other hand, the formalism of quantum states and their dynamics, resources for computational and information processing, are highly non-trivial among non-signaling theories in the sense that entanglement is contained [7] [8]. Contrast to the case of quantum probabilities, capabilities of quantum states have been *tightly* constrained and guided by the no-signaling principle. Namely, assuming entanglement of two parties, the other party's better cloning of quantum states [9] or more efficient quantum dynamics [10] than what can be described within quantum theory are excluded by the no-signaling condition: by the course of such better performances, better *distinguishability* that would establish a faster-than-light communication can be achieved [11].

From the afore-mentioned results, we identify the distinguishability, i.e. distinguishing systems in different states, as the crucial task that derives limitations to other operational tasks. Note also that the distinguishability is generally a fundamental building block for system identifications and practical applications e.g. it determines how efficiently information can be transferred [12], or provides quantification of the universally composable security in cryptographic applications e.g. Refs. [13] [14].

The present work is motivated to consider the underlying state space in which the distinguishability is tightly

constrained by the no-signaling principle: Is it uniquely the space of quantum states? If not, how general is it that the distinguishability is determined generically by the no-signaling principle? To this end, we consider the distinguishability in generalized probabilistic theories (GPTs) [15–17] where states, measurement, and their relations are extended to cover other non-signaling theories. We first develop a theoretical tool for the purpose, a general method of distinguishing states with minimal errors in GPTs. With the established method, we show that the distinguishability is a quantity independent to non-locality: the distinguishability is not enhanced as a theory contains more, or less, non-locality in its bipartite extension. Then, we show that for GPTs where the ensemble steering is allowed, the distinguishability is immediately determined by the no-signaling constraint, regardless of how their state spaces are structured. This shows that the distinguishability in GPTs is generally found as a consequence that ensemble steering on states and the no-signaling condition on probabilities are compromised.

We first begin with the framework of GPTs [15], exploiting the mathematical formalism shown in Ref. [16], see also Ref. [17]. The set of states, denoted by Ω , consists of all possible states that a system can be prepared in. Any probabilistic mixture of states, i.e. $pw_1 + (1-p)w_2 \in \Omega$ for $w_1, w_2 \in \Omega$ and probability p is also a state, and thus the set is convex. A general mapping from states to probabilities is described by *effects*, linear functionals $\Omega \rightarrow [0, 1]$: a measurement s is denoted by a set of effects, $E^{(s)} = \{e_x^{(s)}\}_{x=1}^N$, and the probability of getting outcome x for measurement s when state w is given is, $p(x|s) = e_x^{(s)}[w]$. A unit effect u denotes that a measurement is certain to occur, that is, $u[w] = 1$ for all $w \in \Omega$, so that for any measurement s , it holds $\sum_x e_x^{(s)} = u$. As effects are dual to the state space, they are also convex.

The distinguishability of states in GPT can be described as a game of two parties, Alice and Bob. They have agreed a set of states in advance, and Alice prepares

a system in one of N states with some probability and gives it to Bob. If Bob makes a correct guess, their score is given 1, otherwise 0. The goal is then to maximize the average score over possible measurements, which defines minimum-error state discrimination.

Suppose that states $\{w_x\}_{x=1}^N$ and prior probabilities $\{q_x\}_{x=1}^N$ that Alice generates are in agreement, which altogether are written by $\{q_x, w_x\}_{x=1}^N$. Bob has to find optimal measurement $\{e_x\}_{x=1}^N$ fulfilling $\sum_x e_x = u$, such that he makes guesses from each effect e_x . Let $p_{B|A}(x|y) = e_x[w_y]$ denote the probability that Bob makes a guess w_x when state w_y is given by Alice. The task is to maximize the *guessing probability* in the following,

$$p_{\text{guess}} := \max \sum_{x=1}^N q_x p_{B|A}(x|x) = \max \sum_{x=1}^N q_x e_x[w_x] \quad (1)$$

where the maximization runs over all effects. Note that GPTs are generally not self-dual, meaning an isomorphism between two spaces does not exist in general [18], and thus state and effect spaces are generally distinct.

The sole fact that state and effect spaces are convex allows us to formalize the discrimination problem in the convex optimization framework [19]. This in fact provides a general approach of finding optimal discrimination in GPTs. For states $\{q_x, w_x\}_{x=1}^N$, we take the form in Eq. (1) as the primal problem denoted by p^* and derive its dual d^* , as follows,

$$p^* = \max \left\{ \sum_{x=1}^N q_x e_x[w_x] \mid e_x \geq 0 \forall x, \sum_{x=1}^N e_x = u \right\} \quad (2)$$

$$d^* = \min \{ u[K] \mid K \geq q_x w_x, x = 1, \dots, N \} \quad (3)$$

where inequalities mean the order relation in the convex set: by $e_x \geq 0$, it is meant that $e_x[w] \geq 0$ for all $w \in \Omega$, and by $K \geq q_x w_x$, that $e[K - q_x w_x] \geq 0$ for all effects e .

The property called the strong duality holds true in the above, meaning that both solutions from primal and dual problems are equal, i.e. $p^* = d^*$. This is obtained from the so-called Slater's constraint quantification in convex optimization: a sufficient condition for the strong duality can be the strict feasibility, that is, the existence of a strictly feasible point of parameters. For instance, primal parameters $\{e_x = u/N\}_{x=1}^N$ are in the case, since $e_x[w_y] > 0 \forall x, y$ and $\sum_x e_x = 1$. From this, it is shown that the guessing probability can be obtained from either the primal or the dual problem.

In another approach in convex optimization, called *complementarity problem*, optimality conditions of a given optimization problem are directly analyzed. This deals with both primal and dual parameters in Eqs. (2) and (3), and therefore is not considered more efficient in numerics. The advantage then lies at the fact that generic structures existing in the problem are exploited. The optimality conditions can be summarized by the so-called Karush-Kuhn-Tucker (KKT) conditions. They are

constraints listed in Eqs. (2) and (3), together with the followings,

$$(\text{Symmetry parameter}) \quad K = q_x w_x + r_x d_x, \forall x \quad (4)$$

$$(\text{Orthogonality}) \quad e_x[r_x d_x] = 0, \forall x, \quad (5)$$

where $r_x \in [0, 1]$ for all x , and $\{d_x\}_{x=1}^N$ which we call complementary states are normalized, i.e. $u[d_x] = 1$. The first condition, symmetry parameter, follows from the Lagrangian stability and shows that for any discrimination problem e.g. $\{q_x, w_x\}_{x=1}^N$, there exists a single parameter K which is decomposed into N different ways with given states and complementary states $\{r_x, d_x\}_{x=1}^N$. Then, the second condition in Eq. (5) from the complementary slackness characterizes optimal effects by the orthogonality relation between complementary states and optimal effects. These generalize optimality conditions for quantum states in Refs. [20] [21] to all GPTs, see also different forms of optimality conditions [22].

Primal and dual parameters satisfying the KKT conditions are automatically optimal parameters by which solutions are obtained in the primal and the dual problems. Moreover, for the problem that we consider here, they give the same solution to both problems as the strong duality holds. Conversely, the fact that the strong duality holds in Eqs. (2) and (3) implies the existence of optimal parameters which satisfy KKT conditions and give the guessing probability in Eq. (1). All these follow from the property that state spaces in GPTs forms convex.

Collecting all these, we now formalize a geometric method of solving minimum-error discrimination in GPTs. We first remark that, in optimality conditions in Eqs. (4) and (5), constraints for states and effects are separated. The symmetry parameter K is characterized on a state space and gives the guessing probability, see Eq. (3), i.e. $p_{\text{guess}} = u[K] = q_x + r_x$. Then, the guessing probability can be found by searching complementary states $\{r_x, d_x\}_{x=1}^N$ fulfilling Eq. (4) on the state space. This can be described in a systematic way, as follows. Let us define a polytope denoted by $\mathcal{P}(\{q_x, w_x\}_{x=1}^N)$ of given states in the state space: each vertex of the polytope corresponds to unnormalized state $q_x w_x$ for $x = 1, \dots, N$. Then, the polytope of complementary states, $\mathcal{P}(\{r_x, d_x\}_{x=1}^N)$, is in fact immediately congruent to $\mathcal{P}(\{q_x, w_x\}_{x=1}^N)$ in the state space: from Eq. (4) the following holds,

$$q_x w_x - q_y w_y = r_y d_y - r_x d_x, \text{ for all } x, y, \quad (6)$$

which shows that corresponding lines of two polytopes $\mathcal{P}(\{q_x, w_x\}_{x=1}^N)$ and $\mathcal{P}(\{r_x, d_x\}_{x=1}^N)$ are of equal lengths and anti-parallel. Then, from the underlying geometry of the state space, one can find complementary states by putting two congruent polytopes such that the condition in Eq. (4) holds. Optimal effects can be found from the orthogonal relation in Eq. (5), accordingly.

When prior probabilities are equal i.e. $q_x = 1/N$ for all x , the guessing probability is in a simpler form. First, we have $r_x = r_y$ for all x, y : this is obtained from the expression $p_{\text{guess}} = q_x + r_x$ for any x . Denoted by $r := r_x$ for all x , the guessing probability is characterized as,

$$p_{\text{guess}} = \frac{1}{N} + r, \quad \text{with } r = \frac{\|\frac{1}{N}w_x - \frac{1}{N}w_y\|}{\|d_x - d_y\|} \quad (7)$$

where the expression of r follows from the condition in Eq. (6) with a distance measure $\|\cdot\|$ that can be defined in the state space. The parameter r has a meaning as the ratio between two polytopes, $\mathcal{P}(\{1/N, w_x\}_{x=1}^N)$ of given states, and $\mathcal{P}(\{d_x\}_{x=1}^N)$ of only complementary states.

We illustrate the general method of state discrimination in GPTs, with an example called the polygon systems shown in Refs. [17] [18]. We consider the case of four states, which is of particular interest as its bipartite non-signaling extension can show the maximally non-local correlations. Four states $\{w_x\}_{x=1}^4$ and measurement $\{E^{(x)}\}_{x=1}^4$ with $E^{(x)} = \{e_0^{(x)}, e_1^{(x)}\}$ are given as

$$w_x = \begin{pmatrix} \frac{\cos 2\pi x/4}{\sqrt{\cos \pi/4}} \\ \frac{\sin 2\pi x/4}{\sqrt{\cos \pi/4}} \\ 1 \end{pmatrix}, \quad e_0^{(x)} = \frac{1}{2} \begin{pmatrix} \frac{\cos(2x-1)\pi/4}{\sqrt{\cos \pi/4}} \\ \frac{\sin(2x-1)\pi/4}{\sqrt{\cos \pi/4}} \\ 1 \end{pmatrix}, \quad (8)$$

and $e_1^{(x)} = u - e_0^{(x)}$, where the unit effect $u = (0, 0, 1)^T$, with the Euclidean inner product for $p(a|x) = e_a^{(x)}[w]$.

For four states $\{1/4, w_x\}_{x=1}^4$, the goal is now to find the guessing probability and optimal measurement $\{e_x\}_{x=1}^4$. Exploiting the underlying geometry (cf. see Fig 2. in Ref. [18]), the polytope $\mathcal{P}(\{1/4, w_x\}_{x=1}^4)$ forms a rectangle, from which it follows that $r = 1/4$ from Eq. (7). To be precise, from the state space geometry, one can see that

$$K = \frac{1}{4}w_x + \frac{1}{4}w_{x+2}, \quad \text{for } x = 1, 2, 3, 4, \text{ mod } 4, \quad (9)$$

and thus, $p_{\text{guess}} = 1/2$. It is also shown that complementary states are $\{r_x = 1/4, d_x = w_{x+2}\}_{x=1}^4$. Note that these four states are analogous to cases in quantum theory, pairs of orthogonal states: for the four quantum states, the guessing probability is also given by $1/2$ [22]. This shows that *distinguishability in GPTs may not be enhanced in the way that more non-locality is contained in non-signaling theories*.

Optimal measurement can be found as follows. Putting $e_x = e_0^x$ in Eq. (8) for $x = 1, \dots, 4$, the followings give the guessing probability: i) $\{e_x/2\}_{x=1}^4$, ii) $\{e_1, e_3\}$, or iii) $\{e_2, e_4\}$. In the case ii), effect on e_1 (e_3) means that given state is either w_1 or w_4 (w_2 or w_3), since $e_1[w_2] = e_1[w_3] = 0$. Once effect on e_1 (e_3) is given, one randomly guesses either w_1 or w_4 (w_2 or w_3), and the guessing probability is obtained $1/2$. The case iii) works in a similar way. This shows that features in quantum cases are shared among GPTs: i) optimal measurement is

generally not unique [12], and ii) no-measurement sometimes gives an optimal strategy [22] [23].

We now move to bipartite extensions in GPTs. The extension is specified in an operational way that the ensemble steering is allowed, i.e. in a certain way, any decomposition of Bob's ensemble can be steered by Alice. In quantum theory, this was firstly asserted by Schrödinger [24] and then, with specification to a bipartite Hilbert space, formalized as the so-called Gisin-Hughston-Jozsa-Wootters theorem [25]. Note that, however, GPTs endowed with the ensemble steering do not yet single out quantum theory [26]. We also distinguish the extension from the purification lemma which can characterize quantum theory [27].

In what follows, applying the theoretical tool developed so far for state discrimination, we show that for any GPTs endowed with the ensemble steering, no matter how the state space is structured, the optimal distinguishability is immediately determined by a way of excluding superluminal communication. In other words, in such theories, state discrimination that works better than what is predicted within a given theory itself, would contradict to the no-signaling principle. We first show a bound to the optimal state discrimination from the non-signaling condition, and then prove that the bound is indeed tight i.e. which can be achieved within GPTs.

We first put the task of state discrimination into a non-signaling framework: Bob's distinguishing different decompositions of an identical ensemble steered by Alice at a distance. For states $\{q_x, w_x\}_{x=1}^N$ to discriminate among, suppose that Alice is allowed to steer an ensemble of Bob in N different decompositions, $w_B := w_B^x$ for $x = 1, \dots, N$

$$w_B^{(x)} = p_x w_x + (1 - p_x) c_x, \quad \text{with } q_x = \frac{p_x}{\sum_{x'=1}^N p_{x'}} \quad (10)$$

with some states $\{c_x\}_{x=1}^N$ and probabilities $\{p_x\}_{x=1}^N$. By the ensemble steering, it is meant that any decomposition of an ensemble can be prepared by Alice. Since it is an identical ensemble, Bob's measurement gains no knowledge about which decomposition is given to him until Alice announces about her steering. In this way, a non-signaling scenario is naturally defined.

Let us briefly sketch how to constrain state discrimination using the non-signaling scenario. Suppose that Bob, nevertheless, attempts to distinguish decompositions, to guess about Alice's steering. The strategy exploits discrimination among those states $\{w_x\}_{x=1}^N$ in the ensemble: optimizing measurement for the discrimination and then getting outcome x , he finds that corresponding state w_x exists in the ensemble and concludes that Alice has steered the decomposition w_B^x , see Eq. (10). In this way, state discrimination among states $\{w_x\}_{x=1}^N$ in GPTs cannot work arbitrarily well, and consequently the guessing probability must be limited.

An upper bound to the guessing probability is explicitly derived as follows. First, the no-signaling condition applies to Bob's guessing on average about Alice's steering. Let $P_{B|A}(x|y)$ denote the probability of Bob's concluding x about Alice's steering y after any strategies of him, for which it holds that

$$\sum_{x=1}^N P_{B|A}(x|x) \leq 1, \quad (11)$$

from the no-signaling condition [17] [28], see also Appendix II. If this is not fulfilled, one can explicitly construct a superluminal communication protocol [11]. Next, the strategy of Bob's guessing Alice's steering optimizes measurement for the discrimination among states $\{w_x\}_{x=1}^N$, so that outcome x implies existence of state w_x in the ensemble from which Alice's steering is concluded as the ensemble $w_B^{(x)}$. If Alice has steered Bob's ensemble w_B^x in Eq. (10), Bob's correct conclusion happens when i) w_x is given, which appears with probability p_x , and ii) measurement gives a correct answer: this is with probability $p_x P_{B|A}(x|x)$. In the strategy, there can be contribution in measurement from the other state c_x in the ensemble with probability $1 - p_x$. Thus, it holds, $p_x P_{B|A}(x|x) \leq P_{B|A}(x|x)$. In addition, recall that the discrimination is optimized for $\{q_x, w_x\}_{x=1}^N$, as the *a priori* probability for w_x among states $\{w_x\}_{x=1}^N$ is shown in Eq. (10). From the no-signaling condition in Eq. (11), we have $\sum_{x=1}^N p_x P_{B|A}(x|x) \leq 1$, from which we have

$$p_{\text{guess}} = \max_x \sum_x q_x P_{B|A}(x|x) \leq \frac{1}{p_1 + \dots + p_N}. \quad (12)$$

Thus, a non-signaling bound to the guessing probability in GPTs is obtained with ensemble-steering parameters.

The bound is indeed tight, i.e. which can be achieved within a given GPT. This can be shown by proving that, for any set of states $\{q_x, w_x\}_{x=1}^N$, optimal discrimination implies that both an identical ensemble in Eq. (10) and effects achieving the bound in Eq. (12) exist.

We first recall the existence of a symmetry parameter K that gives the complete characterization of optimal discrimination, see Eq. (4). The parameter has N decompositions with complementary states $\{r_x, d_x\}_{x=1}^N$. Its normalization $\tilde{K} = K/u[K]$ is then, for each x ,

$$\tilde{K} = p_x w_x + (1 - p_x) d_x, \text{ with } p_x = q_x/u[K]. \quad (13)$$

This corresponds to the ensemble steering in Eq. (10). Recall the dual problem in Eq. (3) which gives the guessing probability in GPTs, as $p_{\text{guess}} = u[K]$. Using the identity $\sum_{x=1}^N q_x = 1$ and the relation $p_x = q_x/u[K]$, the solution in the dual problem can be computed as, $u(K) = (\sum_{x=1}^N p_x)^{-1}$. This shows that the bound in Eq. (12) can be achieved within given GPTs, and hence the tightness is shown. In addition, optimal effects are characterized with complementary states, see Eq. (5).

In conclusion, we have developed and established a general method of finding optimal discrimination among states in GPTs, exploiting underlying geometry given in state spaces. This generalizes i) the geometric formulation of quantum state discrimination [22], and ii) optimality conditions in quantum cases [20] [21]. The method is then illustrated in a GPT, with the four-state polygon system shown in Refs. [17] [18]. This is of particular interest as its bipartite extension can achieve the maximally non-local correlations [3]. Applying the general method of optimal discrimination to the example, it is shown that the guessing probability is equal to its analogous case in quantum theory i.e. discrimination of pairs of orthogonal states [22]. This clarifies that operational tasks based on the distinguishability and the non-locality are independent resources: the distinguishability may not be enhanced or decreased as GPTs contain more non-local correlations. From the example, we also remark that for state discrimination in GPTs, i) optimal measurement is generally not unique and ii) no-measurement sometimes give an optimal strategy. All these have been known previously in quantum cases [12] [23].

Then, we have shown that, for GPTs where the ensemble steering is allowed, no matter how their state spaces are constructed or structured, the distinguishability is determined by the no-signaling constraint. This is shown by proving that the upper bound derived by the no-signaling condition can always be achieved within given GPTs. For such theories, the distinguishability lies at the border in which ensemble steering on states and the no-signaling condition on their probabilities are compromised. This means that the tight relation between the no-signaling principle and operational tasks is a feature existing not only in quantum theory but, generally in non-signaling theories.

Finally, discrimination of states is a task applying a single measurement setting, and the present work shows that non-signaling probabilities for optimal discrimination cannot dictate that the underlying state space is quantum-mechanical. If, in some other way, quantum states are found to be unique to describe systems among other non-signaling theories, it is likely that more measurement settings are applied, although non-locality tests themselves are not tightly related to the no-signaling condition. We here leave it open to find such a task - this would be the case that, as unique description of systems, quantum states are the physical entity behind an information processing, together with dimension witnesses [29]. The presented method of finding optimal discrimination in GPTs is a useful theoretical tool to investigate information capabilities and properties of GPTs, e.g., using models shown in Ref. [18].

The author thanks G. Chiribella, D. Gross, M. Navascués, S. Wehner, and S. Yun for helpful discussions and comments. This work is supported by National Research Foundation and Ministry of Education, Singapore.

-
- [1] J. S. Bell, *Physics* **1** 195 (1964).
 - [2] A. Acín *et al.* *Phys. Rev. Lett.* **98**, 230501 (2007).
 - [3] S. Popescu and D. Rohrlich, *Found. Phys.* **24**, 379 (1994).
 - [4] M. Pawłowski *et al.*, *Nature* **461**, 1101 (2009).
 - [5] J. Allcock *et al.*, *Phys. Rev. A* **80**, 040103(R) (2009).
 - [6] R. Gallego, L. E. Wrflinger, A. Acín, and M. Navascués, *Phys. Rev. Lett.* **107**, 210403 (2011).
 - [7] A. J. Short and J. Barrett, *New J. Phys.* **12**, 033034 (2010).
 - [8] D. Gross, M. Müller, R. Colbeck, and O. Dahlsten, *Phys. Rev. Lett.* **104**, 080402 (2010).
 - [9] N. Gisin, *Phys. Lett. A* **242**, 1 (1998).
 - [10] C. Simon, V. Buzek, and N. Gisin, *Phys. Rev. Lett.* **87**, 170405 (2001).
 - [11] J. Bae, W. -Y. Hwang, and Y. -D. Han, *Phys. Rev. Lett.* **107**, 170403 (2011).
 - [12] C. Helstrom, *Quantum Detection and Estimation Theory* Academic, New York, ADDRESS, (1976).
 - [13] R. Canetti, *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science (FOCS-01)*, 136145 (2001).
 - [14] M. Ben-Or, M. Horodecki, D. W. Leung, D. Mayers, and J. Oppenheim, *Proc. of TCC 2005*, LNCS, Vol. 3378, 386406 (2005).
 - [15] G. Birkhoff and J. von Neumann, *Ann. Math.* **37**, 823 (1936).
 - [16] H. Barnum, J. Barrett, and A. Wilce, *Phys. Rev. Lett.* **99**, 240501 (2007).
 - [17] J. Barrett, *Phys. Rev. A* **75** 032304 (2007).
 - [18] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, *New J. Phys.* **13**, 063024 (2011).
 - [19] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, (2004).
 - [20] A. S. Holevo, *Probl. Inf. Transf.* **10**, 317 (1974).
 - [21] H. P. Yuen, R. S. Kennedy, and M. Lax, *IEEE Trans. Inf. Theory* **21**, 125 (1975).
 - [22] J. Bae and W. -Y. Hwang, *arXiv:1204.2313*; J. Bae, *arXiv:1210.2845*.
 - [23] K. Hunter, *Phys. Rev. A* **68**, 012306 (2003).
 - [24] E. Schrödinger, *Proc. Camb. Phil. Soc.* **31**, 553 (1935), *ibid.* **32**, 446 (1936).
 - [25] N. Gisin, *Helv. Phys. Acta* **62**, 363 (1989); L. P. Hughston, R. Jozsa, and W. K. Wootters, *Phys. Lett. A* **183**, 14 (1993).
 - [26] H. Barnum, C. P. Gaebler, and A. Wilce, *arXiv:0912.5532*.
 - [27] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Phys. Rev. A* **84**, 012311 (2011).
 - [28] Ll. Masanes, A. Acín, and N. Gisin, *Phys. Rev. A* **73**, 012112 (2006).
 - [29] N. Brunner *et al.* *Phys. Rev. Lett.* **100**, 210503 (2008).

Appendix I. Convex optimization framework

We show the derivation of primal and dual problems for optimal state discrimination in GPTs. Recall that both state and effect spaces are convex. For convenience, we follow the formalism and notations in Ref. [19]. The

problem to maximize the success probability can be written as the primal problem as,

$$\begin{aligned} \min \quad & f(\{e_x\}_{x=1}^N) = - \sum_{i=1}^N q_x e_x[w_x] \\ \text{subject to} \quad & e_x \geq 0 \quad \forall x, \quad \sum_x e_x = u, \end{aligned}$$

where by $e_i \geq 0$ it is meant that $e_x[w] \geq 0$ for all $w \in \Omega$. Note that the above problem is feasible as the set of parameters satisfying constraints is not empty. It is also strictly feasible with parameters with $e_x = u/N$ for all x .

To derive the dual problem, the Lagrangian can be constructed as,

$$\begin{aligned} \mathcal{L}(\{e_x\}_{x=1}^N, \{d_x\}_{x=1}^N, K) = & f(\{e_x\}_{x=1}^N) - \sum_x e_x[d_x] \\ & + (\sum_x e_x - u)[K], \end{aligned}$$

with $\{d_x \geq 0\}_{x=1}^N$ and K are dual parameters. The dual problem is then obtained by solving the following,

$$g(\{d_x\}_{x=1}^N, K) = \min_{e_x} \mathcal{L}(\{e_x\}_{x=1}^N, \{d_x\}_{x=1}^N, K), \quad (14)$$

for which the Lagrangian can be further evaluated as,

$$\mathcal{L}(\{e_x\}_{x=1}^N, \{d_x\}_{x=1}^N, K) = \sum_x e_x[K - q_x w_x - d_x] - u[K].$$

Then, the minimization in Eq. (14) is to be, $-u[K]$ if $K = q_x w_x + d_x$ for all x , otherwise $-\infty$. Thus, we have $d_x = K - q_x w_x$ for each x . Since $d_x \geq 0$ i.e. $e[d_x] \geq 0$ for all effects e , we write this by, $K \geq q_x w_x$ for each x . The dual problem is therefore as follows.

$$\begin{aligned} \max \quad & -u(K) \quad (\text{or, } \min u[K]) \\ \text{subject to} \quad & K \geq q_x w_x \quad \forall x. \end{aligned}$$

The inequality means an order relation in a convex cone, which is determined by effects, i.e. $e[K - q_x w_x] \geq 0$ for all effects e . Note also that the dual problem is also strictly feasible, with $K = \sum_x q_x w_x$.

Appendix II. No-signaling condition in Eq. (11)

We recall the non-signaling scenario exploited when limiting the guessing probability. It is supposed that Alice can steer Bob's ensemble with N decompositions, labeled by $x = 1, \dots, N$. This does not modify the ensemble average itself of Bob, but only its decompositions. Bob therefore gains no knowledge by all means about which decomposition is given, unless Alice announces her choice about the ensemble steering. In what follows, we explain the no-signaling condition in Eq. (11) in detail.

Let $P_{AB}(a, b|\mathcal{A}, \mathcal{B})$ denote the joint probability distribution of Alice and Bob. To relate this to the scenario, we note that \mathcal{A} corresponds to Alice's choice in the steering, to prepare the decomposition $w_B^{(\mathcal{A})}$ in Eq. (10) for Bob. Then, Bob performs measurement according to her choice \mathcal{B} and makes a guess about the steered decomposition from outcome b . Recall the no-signaling condition on the joint probability shown in Refs. [17] [28], as follows, for all $x \neq x'$

$$\sum_a P_{AB}(a, b|\mathcal{A} = x, \mathcal{B}) = \sum_a P_{AB}(a, b|\mathcal{A} = x', \mathcal{B}). \quad (15)$$

This means that Alice's choice of steering \mathcal{A} does not affect to Bob's statistics, i.e. Bob's measurement-and-outcome relations. This is clear from the ensemble interpretation followed from the assumption of ensemble steering, together with what we have assumed that Bob holds an identical ensemble whose decompositions are steered.

We derive the condition in Eq. (11) by contradiction. For convenience, we write Bob's probability for Alice's steering as, $P_{B|A}(b|\mathcal{A}, \mathcal{B}) = \sum_a P_{AB}(a, b|\mathcal{A}, \mathcal{B})$. With

this, the no-signaling condition in Eq. (15) is expressed as, $P_{B|A}(b|\mathcal{A}, \mathcal{B}) = P_{B|A}(b|\mathcal{A}', \mathcal{B})$. Suppose that for optimal measurement \mathcal{B} , it holds that

$$\sum_{x=1}^N P_{B|A}(b = x|\mathcal{A} = x, \mathcal{B}) > 1 \quad (16)$$

Note that generally Bob's measurement may not be complete: $\sum_{x=1}^N P_{B|A}(b = x|\mathcal{A} = x', \mathcal{B}) \leq 1$ for all x' . Thus, we arrive at the following:

$$\begin{aligned} & \sum_{x=1}^N (P_{B|A}(x|x, \mathcal{B}) - P_{B|A}(x|x', \mathcal{B})) > 0 \\ \implies & \exists x, x' \text{ s.t. } P_{B|A}(x|x, \mathcal{B}) > P_{B|A}(x|x', \mathcal{B}), \end{aligned} \quad (17)$$

which contradicts to the no-signaling condition in Eq. (15). Or, with probabilities in Eq. (17), one can also construct a faster-than-light communication protocol [11]. Thus, the no-signaling condition in Eq. (11) is shown.